UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

ANALYSIS OF MULTITEMPORAL REMOTE-SENSING IMAGES FOR CHANGE DETECTION: BAYESIAN THRESHOLDING APPROACHES

Lorenzo Bruzzone and Roberto Cossu

2002

Technical Report # DIT-02-0031

Also: to be published in Geospatial Pattern Recognition, Research Signpost/Transworld Research, India, Chapter 15

Analysis of Multitemporal Remote-Sensing Images for Change Detection: Bayesian Thresholding Approaches

Lorenzo Bruzzone and Roberto Cossu

Dept. of Information and Communication Technologies

University of Trento, Via Sommarive, 14, 38050 Povo, Trento, Italy

Telephone: +39-0461-882056; e-mail: lorenzo.bruzzone@ing.unitn.it

ABSTRACT

The recent technological developments in remote-sensing sensors and satellites (e.g., the increased spatial and spectral resolutions of sensors, the increased rivisitation time of satellites) offer the possibility of addressing new applications related to environmental monitoring and natural-resource management. In particular, applications connected with the analysis of multitemporal remote-sensing images are becoming more and more important, also in relation to the increased awareness of politicians of the necessity for a regular and efficient control of the environment. This chapter deals with a key issue in multitemporal data analysis, namely, the automatic detection of changes in pairs of images acquired in the same geographical area at different times. In particular, unsupervised change-detection methods (i.e., methods that do not require any ground-truth information for detecting land-cover changes) are considered. Attention is focused on one of the crucial problems that should be solved to develop fully automatic and unsupervised change-detection approaches, i.e., the selection of the decision threshold to be used to distinguish between changed and unchanged areas. Several methods proposed in the literature are described and compared with approaches recently developed by the authors. In addition, some theoretically well-founded thresholding strategies, generally not used for remote-sensing problems, are considered and evaluated also in relation to applications. Examples of results obtained by applying the presented techniques to real multitemporal remote-sensing data sets are reported and discussed.

1. INTRODUCTION

The development of effective methodologies and efficient tools for the analysis of multitemporal remote-sensing images represents one of the most important challenges that the remote-sensing community should face in the next years. The availability of an ever increasing number of remote-sensing satellites characterized by a short rivisitation time makes it possible to consider new advanced applications of remote sensing, ranging from environmental control to monitoring of urban areas. The improvements in the features of remote-sensing sensors, namely, the increased spatial and spectral resolutions, in relation to the temporal dimension, offer the possibility of addressing issues that have not been considered until now with satellite data. However, all scientists and end-users familiar with remote sensing can easily realize the complexity involved in the analysis of hyperspectral series of remote-sensing data or in the processing of multitemporal images characterized by high spatial resolution. From this perspective, the methods currently available for the analysis of multitemporal remote-sensing images do not seem effective enough to support and exploit the technological improvements in satellite and sensor characteristics.

From a general point of view, the terms “analysis of multitemporal remote-sensing images” concern several different applications, which in turn involve using different sensors and methodologies. Applications range from monitoring and management of natural resources (e.g., forests, sea, etc.) to monitoring of land-cover dynamics (e.g., monitoring of ecosystems, monitoring of agricultural areas, etc.), from risk assessment (e.g., forest fires, landslides, floods, etc.) to damage mapping (e.g., burned areas, flooded areas, etc.), from assessment of urban

expansion to updating of road maps on GIS systems. Many of these applications are related to important political issues (e.g., the Kyoto protocol [1]) and play a strategic role for decision makers and politicians.

The large number of possible applications result in different methodological issues in data analysis and processing. Methodologies suited to addressing a problem are often completely different from others adequate to solving another problem. For example, techniques used for the analysis of series of low spatial resolution images acquired in a given forest area are entirely different from approaches to detecting land-cover changes in a pair of high spatial resolution images acquired in a urban area.

In the near future, the pattern recognition and image/signal processing community working on remote-sensing data should devote considerable effort to developing automatic information-extraction methods for translating technological improvements in satellite and sensor characteristics into practical additional services for end-users. This is a mandatory task for proving that remote sensing is an effective technology for resolving problems related to the analysis of multitemporal images.

The above overview points out the large number of issues and open problems dealing with the analysis of multitemporal remote-sensing images, and stresses the topical aspects of this subject. In this chapter, we focus our attention on one of these issues, as it results to be very important in several applications: the problem of detecting land-cover changes by analyzing images acquired in the same geographical area at different times. In particular, unsupervised change-detection techniques (i.e. techniques that do not rely on any ground truth information) based on the analysis of a "difference image" are described. Attention is focused on one of the crucial problems that should be solved to develop fully automatic and unsupervised change-detection approaches i.e., the selection of the decision threshold to be used to distinguish between changed and unchanged areas in the difference image.

The chapter is organized into six sections. Section 2 defines the problem of change detection in multitemporal remote-sensing images and focuses attention on techniques using the difference image. Section 3 deals with heuristic approaches to thresholding the difference image. Section 4 presents some novel techniques recently developed by the authors within a Bayesian framework. In addition, some theoretically well-founded thresholding strategies, generally not used for remote-sensing problems, are described and evaluated. Section 5 gives some examples of applications of the considered techniques to real remote-sensing problems. Finally, in Section 6, results are discussed and conclusions are drawn.

2. UNSUPERVISED CHANGE-DETECTION TECHNIQUES BASED ON THE DIFFERENCE IMAGE

Change detection in multitemporal remote-sensing images is characterized by several peculiar factors that render ineffective some of the multitemporal-image analysis techniques typically used in other application domains. The main difficulties affecting change detection in remote-sensing images arise out of [2], [3], [4]: lack of *a priori* information about the shapes of changed areas; absence of a reference background; differences in light conditions, atmospheric conditions, sensor calibration, and ground moisture at the two acquisition dates considered; problems of alignment of multitemporal images (registration noise). These factors restrict the use of most classical multitemporal image-analysis techniques to few particular remote-sensing problems; for instance, model-based approaches can be adopted only for special purposes, like detection of specific man-made objects [5].

In the literature, several both supervised and unsupervised techniques for detecting changes in multitemporal remote-sensing images have been proposed [2]-[8]. The former require the availability of a "ground truth" from which to derive a training set containing information about the spectral signatures of the changes that occurred in a given area between two dates. The latter perform change detection without any additional information besides the raw images considered. Therefore, from an operational point of view, it is obvious that using unsupervised techniques is mandatory in many remote-sensing applications, as suitable ground-truth information is not always available.

This chapter deals with the widely used type of unsupervised techniques that perform change detection through a direct comparison of the original raw images acquired in the same area at two different times. The change-detection process performed by such techniques is usually divided into three main sequential steps: 1) pre-processing, 2) image comparison, and 3) analysis of the difference image. These steps are detailed in the following.

2.1 Pre-processing

Unsupervised change-detection algorithms usually take two digital images as input and return the locations where differences in the two images can be observed. To accomplish such a task, a preprocessing step is necessary that is aimed at rendering the two images comparable in both the spatial and spectral domains.

Concerning the spatial domain, the two images should be co-registered so that pixels with the same coordinates in the images may be associated with the same ground area. This is a very critical step, which, if inaccurately performed, may make change-detection results unreliable (we refer to [9] for more details on the impact of registration noise on the accuracy of change detection, to [10]-[12] for techniques aimed at supporting the registration process, and to [13]-[15] for techniques devoted to reducing registration noise).

With regard to the spectral domain, changes in light and atmospheric conditions between the two acquisition times may be a potential source of errors and should be taken into account in order to obtain accurate results [16]-[18]. This problem can be mitigated by performing a radiometric calibration of the images. To this end, two different approaches can be taken: absolute calibration and relative calibration. The former involves the conversion of the gray-level values in the images into the corresponding ground reflectance values [19]-[20]. The latter aims at modifying the histograms of the images so that the same gray-level values in the two images may represent the same reflectance values, whatever the reflectance values on the ground may be [16], [17], [21]. The choice of one of the two approaches depends on the particular application considered and on the specific information available.

2.2 Image comparison

The two registered and corrected images are compared, pixel by pixel, in order to generate a further image ("difference image"). The difference image is computed in such a way that pixels associated with land-cover changes present gray-level values significantly different from those of pixels associated with unchanged areas [2]. For example, the univariate image differencing (UID) technique [2], [3] generates the difference image by subtracting, on a pixel-by-pixel basis, a single spectral band from each of the two multispectral images under analysis. The choice of the spectral band to be subtracted depends on the specific type of change to be detected. Other techniques, like vegetation index differencing (VID), make the same kind of comparison by using vegetation indices [18] or a linear (e.g., Tasselled Cap Transformation [18]) or nonlinear combination of original spectral channels. An analogous concept is followed by the widely used change vector analysis (CVA) technique. In this case, several spectral channels are considered at each date (i.e., each pixel of the image considered is represented by a vector whose components are the gray-level values associated with that pixel in the different spectral channels selected). Then, for each pair of corresponding pixels, the so-called "spectral change vector" is computed as the difference in the feature vectors between the two times. The pixels in the difference image are associated with the magnitudes of the spectral change vectors; it follows that unchanged pixels present small gray-level values, whereas changed pixels present rather large values. Another technique similar to the above-described ones is image ratioing (IR) [2]. According to this technique, the comparison between the spectral bands at the two times is accomplished by computing the ratio, instead of the difference, between the two images. The techniques based on the principal component analysis (PCA) can be used for unsupervised change detection in two different ways: by applying the principal component transform separately to the two feature spaces at each single time or to the merged feature space related to the two times. In the first case, the difference image is derived according to the VID technique by using principal components instead of vegetation indices. In the second case, changes result in images associated with the minor components of the transformation.

2.3 Analysis of the difference image

Land-cover changes can be detected by applying a decision threshold to the histogram of the difference image. For instance, when the UID or the CVA techniques are used, changed pixels can be identified on the right side of the histogram, as they are associated with large gray-level values. The selection of the decision threshold is of major importance, as the accuracy of the final change-detection map strongly depends on this choice. This last step is the most critical one in the development of completely automatic and unsupervised techniques for the detection of land-cover changes. Therefore, in the following, we shall focus our attention on the problem of the automatic analysis/thresholding of the difference image.

2.4 Problem formulation

Let us introduce the notation used in the following of the chapter. Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be two multispectral images, of size $I \times J$, acquired in the same geographical area at two different times, t_1 and t_2. Let us assume that such images have been co-registered, and that the possible differences in the light and atmospheric conditions at the two times have been corrected. Let $\boldsymbol{X}_D$ be the difference image obtained by applying to the original images one of the techniques described in subsection 2.2. Finally, let $x(i, j)$ be the pixel with coordinates (i, j) in $\boldsymbol{X}_D$, and let $X(i, j)$ be a random variable (in the range $[0,1,...,G-1]$) that represents the value assumed by $x(i, j)$. Let w_n and w_c be the classes associated with unchanged and changed pixels, respectively. The approaches to the analysis/thresholding of the difference image $\boldsymbol{X}_D$ can be divided into two categories: heuristic approaches and Bayesian approaches. The former include subjective non-automatic strategies based on trial-and-error procedures or image thresholding algorithms aimed at optimizing empirical functions [2], [22], [23]. The latter consist of strategies developed within the framework of the Bayesian decision theory [24], [25].

3. AUTOMATIC ANALYSIS/THRESHOLDING OF THE DIFFERENCE IMAGE: HEURISTIC APPROACHES

In remote-sensing problems, the analysis of the difference image is usually carried out according to empirical thresholding strategies or manual trial-and-error procedures. In particular, two main categories of methods have been

proposed in the literature: i) methods based on statistical models; ii) methods based on the optimization of an objective function of the selected threshold.

3.1 Methods based on statistical models

This category includes methods based on the modeling of: i) the distribution of the difference image; ii) the spatial distribution of noise in the difference image; iii) the spatial distribution of the signal in the difference image.

The methods based on the modeling of the distribution of the difference image are usually developed within the framework of the single-hypothesis testing theory. Such a theory implies the assumption (reasonable but not always verified) that few changes occurred between the two dates considered. Under this hypothesis, the density function of the pixel gray-level values in the difference image may be confused with the density function of the unchanged pixels. Consequently, a decision strategy based on the single-hypothesis testing theory can be adopted [26], i.e., pixels with gray-level values significantly different from the mean of the density function of the difference image are labeled as changed. In many cases, it is reasonable to assume that the distribution of the difference image can be modeled with a Gaussian function. Under this assumption, the decision threshold is typically fixed at $n s_D$ from the mean value m_D of the difference image, s_D being the standard deviation of the density function of the pixel gray-level values in the difference image, and n being a real number derived by a trial-and-error procedure. Some authors [8], [27], [28] have experimentally studied the effects of different n values on the accuracy of change-detection results.

The methods based on the spatial distribution of noise focus attention on the model of the spatial noise affecting the difference image. Usually, this noise (which is assumed to be white) is modeled according to a Poisson distribution. The threshold selection process is carried out on the basis of the observation that pixels belonging to changed regions tend to cluster rather than follow a Poisson distribution [22].

The connectivity method belongs to the category of approaches that model the spatial distribution of the signal in the difference image. The rationale for this method is the following. Considering only the signal and supposing that the classes are well separated, it is reasonable to assume that the properties of the regions made up of changed and unchanged pixels will remain stable over a range of thresholds; on the contrary, considering only the spatial noise, small differences in the threshold value may lead to significant variations in the number and properties of the changed areas detected in the image. On the basis of this observation, O'Gorman [29] proposed first to look for a range of threshold values that lead to a stable number of regions and then to select the final threshold from this range. Rather than counting the number of regions, Rosin and Ellis [30] suggested to use the image's Euler number to increase the computational efficiency of the method. According to this approach, the decision threshold can be set at the corner of the curve that describes the behavior of the aforementioned Euler number versus the value point. Such a corner is detected as the point with the maximum deviation from the straight line drawn between the end points of the curve.

3.2 Methods based on the optimization of an objective function

The methods of this category threshold the difference image in such a way that the changed and unchanged classes satisfy a set of desired properties. This task is accomplished by optimizing (i.e., maximizing or minimizing) an objective function of the selected threshold. These approaches have been derived from the field of image segmentation, where they are widely used. Otsu [31] proposed an interesting technique belonging to this kind of methods. It maximizes the separability of changed and unchanged pixels by considering the variance between the related classes. The Kapur method [32] and the Huang and Wang method [33] maximize the entropy measure of the segmented histogram. The thresholds provided by these entropy criteria tend to make the obtained class means as separated as possible and the class variances as small as possible.

4. AUTOMATIC ANALYSIS/THRESHOLDING OF THE DIFFERENCE IMAGE: BAYESIAN APPROACHES

The weak link of heuristic approaches with the optimal Bayesian decision theory, as well as their dependence on subjective and empirical criteria, represents a critical limitation of this kind of methods. An interesting alternative, aimed at overcoming the aforementioned drawbacks, consists in formulating the problem of the analysis of the difference image in the framework of the Bayesian decision theory. Within this framework, change detection can be defined as a two-class classification problem: we have to discriminate between two classes, w_n and w_c, associated with unchanged and changed pixels, respectively. Given the one-dimensional nature of the problem of the difference-image analysis, the solution of the classification problem is equivalent to the selection of a decision threshold. However, the application of the Bayesian theory requires the knowledge of the statistical terms associated with the two classes (i.e., prior probabilities $P(w_n)$, $P(w_c)$ and probability density functions $p(X/w_n)$, $p(X/w_c)$). This represents a critical issue in the unsupervised analysis of the difference image, as we have not a training set for estimating the aforementioned terms. In [34], a computationally efficient solution to the problem is proposed. The method is based on the optimization of a criterion function related to the average

pixel classification error rate rather than to the estimation of the statistical terms of the changed and unchanged classes. On the one hand, this method turns out to be very simple and computationally efficient; on the other hand, the derived threshold value corresponds to the one that would be obtained by using biased estimates of the required statistical terms. In [24], [25], an unsupervised approach is presented that aims at explicitly estimating the aforesaid statistical terms of the changed and unchanged classes on the basis of the properties of the difference image. Thanks to the resulting estimates, it is then possible to apply different Bayesian strategies to the analysis/thresholding of the difference image. In the following, we focus on this unsupervised approach.

4.1 Unsupervised estimation of the difference-image statistical terms

For the sake of simplicity, let us assume that the difference image is obtained with the UID or the CVA techniques (it is worth noting that a generalization to other techniques is straightforward). The unsupervised estimation of the statistical terms of the difference image is based on the assumption that the histogram of $\boldsymbol{X}_D$ can be modeled as a mixture density made up of the distributions of the two classes associated with the changed and unchanged pixels, respectively, i.e.:

$$p(X) = p(X/w_n)P(w_n) + p(X/w_c)P(w_c). \qquad (1)$$

According to (1), the estimation of these statistical terms can be carried out in the framework of a mixture estimation problem [24], [25]. Two possible approaches can be taken: the parametric approach [24] and the adaptive semiparametric one [25]. The former assumes that the two considered classes can be modeled with a parametric distribution (e.g., a Gaussian distribution). The latter does not assume any specific model for the distribution of the classes, which are described by a "sub-optimal" set of kernel functions. Both estimation approaches include two different phases: i) initialization; ii) iterative optimization.

i) *Initialization* - In this phase, a small number of pixels, which can be reasonably labeled as belonging to either w_n or w_c, are selected by exploiting the intrinsic characteristics of the difference image. Then, on the basis of such pixels, the initial estimates of the statistical terms involved in (1) are derived. In particular, a subset $\boldsymbol{S_n}$ of pixels likely to belong to w_n and a subset $\boldsymbol{S_c}$ of pixels likely to belong to w_c can be obtained by applying two thresholds, T_n and T_c, to the left and right extremes of the histogram $h(X)$ of the difference image (see Fig. 1). Different strategies can be adopted in order to obtain two reasonable values of T_n and T_c, depending on the prior information available for the problem considered [25]. A simple strategy consists in defining T_n and T_c in terms of the middle value M_D of the histogram $h(X)$ as:

$$T_n = M_D(1-a); \qquad T_c = M_D(1+a) \qquad (2$$

where a is a real number ($0 < a < 1$) that defines the range around M_D in which pixels cannot be easily classified as either changed or unchanged pixels. M_D can be expressed by $M_D = (max\{\boldsymbol{X_D}\} - min\{\boldsymbol{X_D}\})/2$, where $max\{\boldsymbol{X_D}\}$ and $min\{\boldsymbol{X_D}\}$ are the maximum and minimum gray-level values in $\boldsymbol{X_D}$, respectively. Once T_n and T_c have been fixed, first rough estimates of $p(X/w_n)$, $p(X/w_c)$, $P(w_n)$ and $P(w_c)$ can be obtained from the initial sets of pixels, $\boldsymbol{S_n}$ and $\boldsymbol{S_c}$.

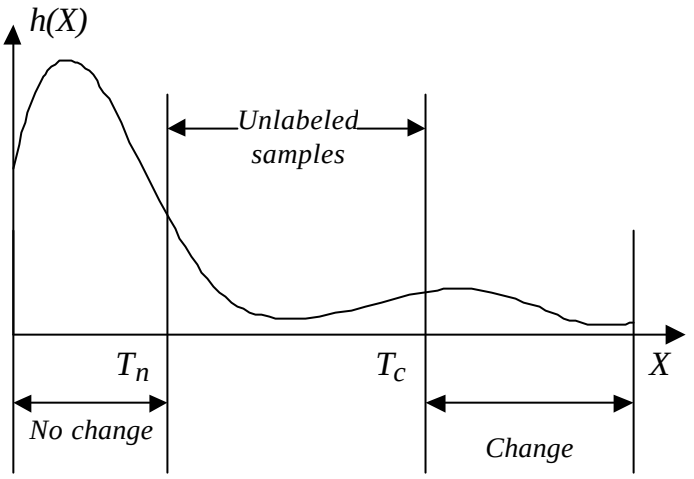


Fig. 1. Schematic representation of the thresholding strategy applied to the difference-image histogram in the initialization phases of the unsupervised estimation approaches.

ii) Iterative optimization - In this phase, also the unlabeled pixels that have not been used in the first phase, and that can turn out to be either changed or unchanged, are used to improve the estimates of the aforesaid statistical terms by an iterative process based on the EM algorithm [35]-[37]. In particular, the subset of pixels $\boldsymbol{S_u} = \boldsymbol{X_D} - \{\boldsymbol{S_n} \cup \boldsymbol{S_c}\}$ is used to increase the accuracies of the estimates of the statistical terms that define the mixture density $p(X)$. The estimates obtained by the EM algorithm at convergence can be exploited to select the threshold for the difference image on the basis of the Bayesian theory (the strategies that can be adopted within this framework are described in subsection 4.2).

Although the general formulation of the methodology is similar, the parametric and the adaptive semi-parametric approaches result in different specific definitions of the aforementioned initialization and iterative optimization steps. In the following, the two approaches are considered separately.

4.1.1 Parametric estimation of $p(X/w_n)$, $p(X/w_c)$, $P(w_n)$, and $P(w_c)$

Let us consider, for simplicity, that the distribution of the classes of changed and unchanged pixels can be modeled by a Gaussian function (it is worth noting that, in this approach, any parametric model could replace the Gaussian one). This assumption is reasonable, for example, if we derive the difference image by applying the UID technique to images acquired by passive sensors. It remains a reasonable approximation also if the CVA technique is applied instead of the UID one (however, note that in this case the magnitude operator may affect the Gaussianity of the density function of unchanged pixels).

Under this assumption, the density function associated with the class w_n can be described by the mean m_n and the variance s_n^2; analogously, the density function associated with the class w_c can be described by the mean m_c and the variance s_c^2. Accordingly, the parameter vector to be optimized is $q = [m_n, m_c, s_n^2, s_c^2]$. This vector is initialized by computing the means and the variances of the classes by using the pixels belonging to the subsets $\boldsymbol{S}_n$ and $\boldsymbol{S}_c$. Then, it is possible to prove that the iterative equations to be used to optimize the components of the vector in an unsupervised way are the following (consider the class w_n) [24]:

$$P^{t+1}(w_n) = \frac{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(w_n)p^t(X(i,j)/w_n)}{p^t(X(i,j))}}{I \cdot J} \quad (3)$$

$$m_n^{t+1} = \frac{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(w_n)p^t(X(i,j)/w_n)}{p^t(X(i,j))} X(i,j)}{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(w_n)p^t(X(i,j)/w_n)}{p^t(X(i,j))}} \quad (4)$$

$$(s_n^2)^{t+1} = \frac{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(w_n)p^t(X(i,j)/w_n)}{p^t(X(i,j))} \left[X(i,j) - m_n^t\right]^2}{\sum_{X(i,j)\in \mathbf{X_D}} \frac{P^t(w_n)p^t(X(i,j)/w_n)}{p^t(X(i,j))}} \quad (5)$$

where the superscripts t and $t+1$ denote the values of the parameters at the current and next iterations, respectively. Analogous equations are used to estimate the prior probability and the mean and variance values of the conditional density function associated with the class w_c. It is possible to prove that, at each iteration of the EM algorithm, the estimated parameters provide an increase in the log-likelihood function $L(q) = \ln p(X \mid q)$. At convergence, a local maximum of this function is reached. Although convergence can be ensured, it is not possible to guarantee that the algorithm will converge to the global maximum of the likelihood [37]-[39].

4.1.2 Adaptive semi-parametric estimation of $p(X/w_n)$, $p(X/w_c)$, $P(w_n)$, and $P(w_c)$

The parametric estimation procedure can be used only when a reasonable assumption on the distribution of the classes can be made. However, in some cases (e.g., when the difference image is computed on the basis of texture features, when the classes are multimodal, etc.), it is not possible to assume any parametric model for the class distributions. In this case, the adaptive semi-parametric approach proposed in [25] should be used. The term "adaptive" points out that this approach does not assume any *a priori* model for the data distribution; the term "semi-parametric" refers to the nature of the final estimates, which are derived by converting the initial non-parametric model into a more suitable semi-parametric description of the difference-image densities. This procedure exploits two well-founded theoretical techniques for accomplishing the estimation process: the reduced parzen estimate (RPE) technique [40] and the EM algorithm. In particular, the RPE procedure is used to derive initial non-parametric estimates of the probability density functions of changed and unchanged pixels in the difference image. Then, these non-parametric estimates are iteratively improved by using the EM algorithm to provide a more accurate description of the difference-image statistics. In this case, the method is significantly more complex that the parametric one.

In the first phase (initialization), the initial estimates of the statistical terms involved in (1) are derived by applying the non-parametric RPE technique [40] separately to the subsets of pixels belonging to $\boldsymbol{S}_n$ and $\boldsymbol{S}_c$. Accordingly, the density function $p(X/w_V)$ ($w_V \in \{w_n, w_c\}$) can be initialized by using the following equation:

$$\hat{p}_{R_V}(X/w_V) = \frac{1}{R_V} \sum_{r=1}^{R_V} \frac{1}{h} k\left[\frac{(X - Y_r)}{h}\right] \quad (6$$

where $k(\cdot)$ is a kernel function satisfying $\int k(X)dX = 1$, h is the kernel-size control parameter, Y_r represents each of the R_n representative pixels selected from $\boldsymbol{S}_n$ ($\boldsymbol{S}_V \in \{\boldsymbol{S}_n, \boldsymbol{S}_c\}$) so that they can approximate the distribution of the pixels in $\boldsymbol{S}_n$ as close as possible [25]. According to [40], the R_n representative pixels are selected from $\boldsymbol{S}_n$ so that the entropy between (6) and the corresponding estimates carried out by the classical Parzen estimation process (applied to all samples) can be maximized [25]. The resulting estimates, even though biased by the two initial sets of pixels, $\boldsymbol{S}_n$ and $\boldsymbol{S}_c$, represent a reasonable starting point to derive the final density functions of changed and unchanged

pixels in the difference image. For a more detailed description of the initialization technique, we refer the reader to [25].

In the second phase (iterative optimization), a more flexible expression for the estimation of each density function $p(X/w_n)$ is applied. In particular, the kernel-size control parameter is allowed to be different for each kernel; in addition, each kernel function is associated with a different weight such that the sum of the R_n weights associated with the class w_n may be equal to one. On the one hand, this increases the number of parameters to be estimated but, on the other hand, it makes it possible to obtain more accurate approximations for the real distributions. It is worth noting that the aforementioned procedure implies the conversion of the non-parametric model (derived from the previous initialization phase) into a more suitable semi-parametric model. This semi-parametric model provides (by exploiting the unlabeled samples present in the subset S_u) a more accurate description of the probability density functions of changed and unchanged pixels. In this context, the new expression for each density function $p(X/w_n)$ can be obtained by rewriting expression (10) as:

$$\hat{p}_{R_v}(X/w_v) = \sum_{r=1}^{R_v} \frac{\Pi_{r,v}}{h_{r,v}} k\left[\frac{(X - Y_{r,v})}{h_{r,v}}\right]. \qquad (7)$$

Using this formulation, the parameter vector to be optimized becomes $q_v = [h_{1,v},...,h_{R_v,v},Y_{1,v},...,Y_{R_v,v},P_{1,v},...,P_{R_v,v}]$, where $h_{l,v}$, $Y_{l,v}$ and $\Pi_{l,v}$ represent the *l-th* kernel-size control parameter, the *l-th* representative sample and the *l-th* weight, respectively, associated with the class w_n.

In this case too, it is possible to formulate the iterative estimation process in the framework of the EM algorithm [25]. At each iteration of the EM algorithm, the estimated parameters provide an increase in the log-likelihood function $L(q) = ln\, p(X \mid q)$ (where $q = [q_n, q_c]$ is the parameter vector to be computed) [25]. At convergence, a local maximum of the log-likelihood function is reached.

4.2 Thresholding strategies

The unsupervised estimation procedures described in the previous subsection allow one to derive the statistical terms necessary to define the problem of selecting the decision threshold in a Bayesian framework. Accordingly, different strategies can be used to derive the threshold value. The most suitable strategy should be chosen on the basis of both end-users' requirements and the available *a priori* knowledge of the considered problem. Concerning end-users' requirements, two different goals can be attained: minimization of the overall change-detection error and minimization of the consequences of errors on the specific application considered. Moreover, in the latter case, it is possible to deal with two different situations, depending on the available prior knowledge: the different consequences involved by false and missed alarms can be quantitatively assessed (this allows one to associate a cost term with each kind of error); quantitative evaluations of the costs of different errors are not possible but a constraint on the percentage of false (or missed) alarms can be imposed. A further distinction should be made between two possible operating conditions: the system to be developed can support the iterative approaches to the estimation of the statistical terms of the difference image described in the previous section; the system can rely on a fixed threshold derived in the system-design phase, but cannot update this value over time by automatic procedures. In the following, these cases will be discussed in greater detail.

4.2.1 Minimum-error strategy

Let us consider the case in which we are interested in minimizing the overall change-detection error without considering the costs of false and missed alarms. The minimization of the overall change-detection error can be accomplished by selecting the decision threshold according to the Bayes rule for the minimum error, i.e. [41]

$$w_k = \arg \max_{w_i \in \{w_n, w_c\}} \{P(w_i / X(i,j))\} = \\ = \arg \max_{w_i \in \{w_n, w_c\}} \{P(w_i) p(X(i,j)/w_i)\} \qquad (8$$

Applying this rule to solve the change-detection problem is equivalent to thresholding the difference image at the maximum-likelihood boundary T_o between the classes w_n and w_c. Therefore, on the basis of the estimates of the statistical terms obtained by the EM algorithm, the optimum threshold value T_o can be estimated by solving the following equation in terms of the variable X:

$$\frac{P(w_c)}{P(w_n)} = \frac{p(X/w_n)}{p(X/w_c)}. \qquad (9$$

It is worth noting that the accuracy of the threshold value $\hat{T}_o$ obtained and hence the precision of the final change-detection map depend on the accuracies of the estimates provided by the EM algorithm.

The use of this strategy results in the implicit assumption that false and missed alarms are of the same importance in the considered application.

4.2.2 Minimum-cost strategy

In several practical applications, to assume that false and missed alarms are of the same importance is not the best strategy to accomplish the change-detection process. Different kinds of errors may be associated with different consequences to end-users. This typically occurs when

change-detection maps are used to support decision making (e.g., in relation to risk assessment, damage evaluation, etc.). From this perspective, the decision threshold should be defined according to a tradeoff between false and missed alarms weighted with the costs they involve in the application considered. If a quantitative evaluation of costs of false and missed alarms is possible (or if at least the relative importance of such costs can be estimated), the decision threshold can be derived according to the Bayes rule for the minimum cost [42], [43]. This rule is a generalization of the Bayes rule for the minimum error.

Let c_{cn} be a positive integer value that defines the cost of false alarms (i.e., the cost of classifying a pixel belonging to the class w_n as a pixel belonging to the class w_c), and let c_{nc} be a positive integer value that defines the cost of missed alarms (i.e., the cost of classifying a pixel belonging to the class w_c as a pixel belonging to the class w_n). The decision threshold T_o that minimizes the cost involved by the decision rule can be derived by solving the following equation in terms of the variable X:

$$\frac{c_{nc}P(w_c)}{c_{cn}P(w_n)} = \frac{p(X/w_n)}{p(X/w_c)} \tag{10}$$

Equation (10) can be rewritten as:

$$\frac{P(w_c)}{kP(w_n)} = \frac{p(X/w_n)}{p(X/w_c)} \tag{11}$$

where $k = c_{nc}/c_{cn}$ is a parameter that simplifies the definition of the threshold as it establishes the relation between the costs of false and missed alarms, instead of assigning numerical values to them [41].

4.2.3 Neyman-Pearson strategy

In some problems, it is very difficult to determine realistic costs of false and missed alarms because it is complex to associate the consequences of each error with an appropriate integer number. To overcome this drawback, it is possible to define the decision threshold by taking into account the conditional probabilities of false alarms (P_f) and missed alarms (P_m), i.e.,

$$P_f = \int_{T_o}^{G-1} p(X/w_n)dX \tag{12}$$

$$P_m = \int_{0}^{T_o} p(X/w_c)dX \; . \tag{13}$$

In particular, we may aim to minimize the overall change-detection error subject to a constraint on P_f or P_m. For example, we might aim to minimize the total error subject to the constraint that the probability of false alarms (or missed alarms) should be fixed at a predefined value. This allows one to deal with applications where an increase in false (or missed) alarms above a given value results in unacceptable consequences to end-users. In these cases, the decision strategy should be designed according to the Neyman-Pearson strategy [41]. In particular, given the one-dimensional nature of our problem, the decision threshold can be easily obtained on the basis of equation (12) (if the probability of false alarms is fixed) or of equation (13) (if the probability of missed alarms is fixed) (see Fig. 2). The density functions $p(X/w_n)$ and $p(X/w_c)$ to be considered for the computation of the decision threshold are the ones estimated by using the approaches described in subsection 4.1.

It is worth noting that, in our case, there is a strong relationship between the Bayes rule for the minimum cost and the Neyman-Pearson criterion. In particular, they are based on a similar philosophy. The only difference is that, if we use the former, we explicitly consider the costs of both false and missed alarms. If we use the latter, we fix the threshold in order to minimize the overall error probability by imposing a constraint on the rate of false alarms (or missed alarms).

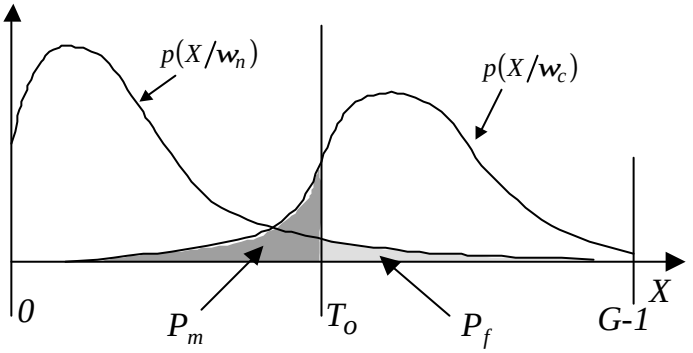


Fig. 2. Probabilities of false alarms (P_f) and missed alarms (P_m) involved by the decision threshold T_o.

4.2.4 Minimax strategy

All the above-described decision rules require the knowledge of the prior probabilities of the classes $P(w_c)$ and $P(w_n)$. In the context of our problem, this is not a critical constraint, given the available unsupervised procedure for the estimation of the prior probabilities and the probability density functions of the classes. However, in some applications, for a specific change-detection problem (e.g., detection of burned areas), it might be necessary to set a decision threshold in the design phase of the system, and then to rely on this threshold also for new images different from the ones used in the threshold-selection phase. Even if this situation should be avoided because the resulting change-detection maps might be seriously affected by this procedure, it is important to use an appropriate strategy (for

the threshold-selection process) capable to deal with this kind of problems. In this context, if the application, the geographical area, and the sensor used do not change, it is quite reasonable to assume that if one changes the acquisition times of the images, the behaviors of the density functions of changed and unchanged areas will not significantly vary. On the contrary, it is evident that the prior probabilities of changed and unchanged classes may result in significant variations over time. Consequently, in the design phase of the system, a "robust" decision rule should be defined, i.e., a threshold should be chosen that is capable to discriminate between changed and unchanged pixels as well as possible over a wide range of different values of the prior probabilities. To deal with this problem, we propose to adopt the minimax criterion [41]. This criterion, which is well-known in classical decision theory, allows one to select the threshold by minimizing the maximum possible overall error (or cost) versus the values of the prior probabilities of the classes. In other words, the minimax criterion defines such a threshold that the worst overall error (or cost) for any values of the prior probabilities turns out to be as small as possible. It can be proven that the equation that should be used to derive the decision threshold according to the minimax \criterion is the following [41]:

$$(c_{cn}-1)\int_{T_o}^{G-1} p(X/w_n)dX + (c_{nc}-1)\int_{0}^{T_0} p(X/w_c)dX = 0 \,. \quad (14)$$

Figure 3 shows a graphical description of the minimax decision rule. The curve represents the trend of the Bayes change-detection error versus the prior probability $P(w_c)$ of changed pixels (it is worth noting that, as $P(w_n)$=1- $P(w_c)$, this diagram completely represents the behavior of the error in the considered change-detection problem). This curve is obtained by selecting the optimal decision threshold (according to the rule described in subsection 4.2.1) for any possible value of $P(w_c)$. For any decision threshold chosen for fixed values of the prior probabilities (e.g., for $P(w_c)$=0.2 in Fig. 3), if the prior probabilities vary (i.e., if, in new images, they are different from the ones related to the images used to design the decision rule), the error probability will change as a linear function of $P(w_c)$ (dashed line in Fig. 3). The maximum of such an error will occur at extreme values of the prior probabilities (e.g., at $P(w_c)$=1 in Fig. 3). To minimize the maximum possible values of this error, we should select such a decision threshold that the maximum error may occur ($P(w_n)$ =0.7 in Fig. 3). As a consequence, we have the advantage that the error will not change as a function of the prior probabilities, as shown in Fig. 3 by the solid horizontal line.

It is important to stress that this decision rule usually involves a much higher change-detection error (or change-detection cost) than the optimal Bayes rule for the minimum error (or the minimum cost). However, if in the considered change-detection system the tuning of the decision threshold over time is not possible, the use of the minimax criterion allows one to avoid a dramatic increase in the change-detection error resulting from possible significant changes in the prior probabilities of the classes present in the new images.

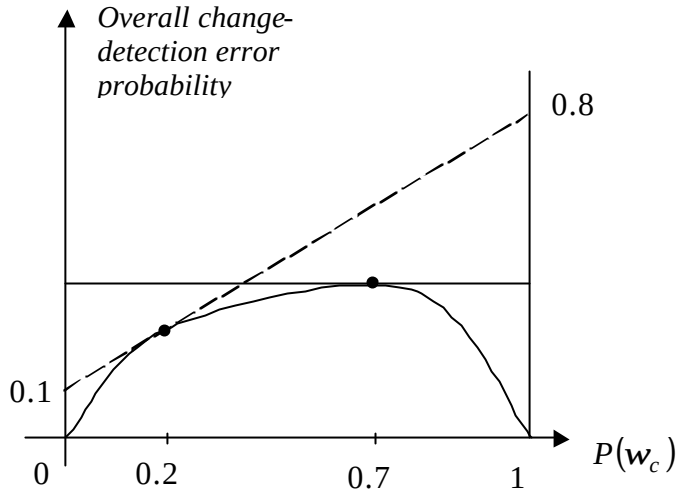


Fig. 3. Graphical description of the minimax decision rule. The curve represents the Bayes change-detection error versus the prior probability $P(w_c)$. Given a decision threshold derived for specific values of the prior probabilities (e.g., $P(w_c)$=0.2), if the prior probabilities vary, the change-detection error changes as a linear function (dashed line). To minimize the maximum vale of such an error, the threshold should be selected for the maximum Bayes error (e.g., $P(w_c)$=0.7). As a result, the overall error will not change as a function of the prior probability (solid horizontal line).

4.2.5. Context-based strategy

The previous described strategies allow one to automatic select the decision threshold value that minimizes the overall change-detection error probability, under the assumption that the pixels in the difference image are independent of one another. An alternative strategy consists in analyzing the difference image by considering the spatial-contextual information included in the neighborhood of each pixel. In [25], an approach based on Markov Random Fields (MRFs) is presented that exploits the interpixel class dependence context (we refer the reader to [44]-[47] for details on MRFs). Such an approach involves the assumption that the changes to be identified are large enough to be detected by the sensor used. Under this hypothesis, a pixel belonging to the class w_k is likely to be surrounded by pixels belonging to the same class. Therefore, efficient use of this interpixel class dependence may yield more reliable and more accurate change-detection results.

Let the set $\mathbf{C} = \{\mathbf{C}_l, 1 \le l \le L\}$, with $L = 2^{IJ}$, be composed of all the possible sets of labels in the difference image $\boldsymbol{X_D}$, where $\mathbf{C}_l = \{C_l(i,j), 1 \le i \le I, 1 \le j \le J\}$, with

$C_l(i,j) \in \{w_n, w_c\}$, is a generic set of labels in $\boldsymbol{X_D}$. By taking into account the spatial-contextual information, the Bayes rule for the minimum error, as defined in (8), can be rewritten as the selection of a set $\mathbf{C}_k$ that maximizes the following rule:

$$\mathbf{C}_k = \underset{\mathbf{C}_l \in \mathbf{C}}{arg\,max} \{P(\mathbf{C}_l / \boldsymbol{X}_D)\} = \\ = \underset{\mathbf{C}_l \in \mathbf{C}}{arg\,max} \{P(\mathbf{C}_l) p(\boldsymbol{X}_D / \mathbf{C}_l)\} \quad (15)$$

where $P(\mathbf{C}_l)$ is the prior model for the class labels, and $p(\boldsymbol{X_D}/\mathbf{C}_l)$ is the joint density function of the pixel values in the difference image, given the set of labels $\mathbf{C}_l$. The maximization of (15) requires the estimations of both $P(\mathbf{C}_l)$ and $p(\boldsymbol{X_D}/\mathbf{C}_l)$, which are very complex tasks. The problem can be simplified if we model the spatial-contextual information in a local spatial neighborhood. This is rather a reasonable approach if we consider the interpixel class dependence, as the interactions between the classes of pixels decrease rapidly as the distances between the pixels increase. In this context, an MRF approach can be used to model the spatial context in the prior model for the class labels $P(\boldsymbol{C}_l)$. MRFs provide a methodological framework that allows one to fully exploit the interpixel class dependence. As a further simplification of the problem, we assume the following conditional independence:

$$p(\boldsymbol{X_D} / \mathbf{C_I}) = \prod_{X(i,j) \in \boldsymbol{X}_D} p(X(i,j)/C_l(i,j)). \quad (16)$$

According to (15), the generation of the final change-detection map involves the labeling of all the pixels in the difference image so that, under the above-mentioned assumptions, the conditional posterior probability is maximized. In terms of the Markovian approach, it is possible to prove [25] that this procedure is equivalent to the minimization of the following energy function [44]-[48]:

$$U(\boldsymbol{X}_D, \mathrm{C}_l) = \sum_{1 \le i \le I} \sum_{1 \le j \le J} [U_{data}(X(i,j)/C_l(i,j)) + \\ + U_{context}(C_l(i,j)/\{C_l(g,h),\ (g,h) \in \mathbf{N}(i,j)\})] \quad (17)$$

where, in the case of Gaussian distributions of the changed and unchanged classes (it is straightforward to derive similar equations for the adaptive semi-parametric case [25]), we have:

- $U_{data}(X(i,j), C_l(i,j)) = \frac{1}{2} \ln \left| 2p s^{\,2}_{C_l(i,j)} \right| + \\ + \frac{1}{2} \left(X(i,j) - m_{C_l(i,j)}\right)^2 \left[s^{\,2}_{C_l(i,j)}\right]^{-1}$ (18)

- $s^{\,2}_{C_l(i,j)} \in \{s^{\,2}_n, s^{\,2}_c\}$ and $m_{C_l(i,j)} \in \{m_n, m_c\}$ are the estimates obtained by the EM algorithm under the assumption of independence.

- $U_{context}(\cdot)$ is the Gibbs energy function given by:
$$U_{context}(C_l(i,j)/\{C_l(g,h),\ (g,h) \in \mathbf{N}(i,j)\}) = \\ = \sum_{(g,h) \in \mathbf{N}(i,j)} b d_k(C_l(i,j), C_l(g,h)) \quad (19)$$

- $\mathrm{N}(i,j) = \{(i,j) + (v,V),\ (v,V) \in \mathrm{N}\}$ is a neighbor of the pixel with coordinates (i,j),

- d_k is the Kronecker delta function given by:
$$d_k(C_l(i,j), C_l(g,h)) = \\ = \begin{cases} -1 & \text{if } C_l(i,j) = C_l(g,h) \\ 0 & \text{if } C_l(i,j) \neq C_l(g,h) \end{cases} \quad (20)$$

- b is a constant that tunes the influence of the spatial-contextual information on the change-detection process.

The minimization of (17) can be carried out by using the simple and fast approach based on Besag's *Iterated Conditional Modes* (ICM) algorithm, which has been proved to converge to a local minimum of the energy function [49]-[50]. According to this approach, the $\mathbf{C}_l$ that minimizes (17) is obtained by the following algorithm:

Step 1: For all pixels $X(i,j) \in \boldsymbol{X_D}$, initialize $C_l(i,j)$ with the class that minimizes the non-contextual energy function $U_{data}(X(i,j), C_l(i,j))$.

Step 2: For all pixels $X(i,j) \in \boldsymbol{X_D}$, update $C_l(i,j)$ to the class that minimizes equation (17).

Step 3: Repeat step 2 until convergence is reached.

5. EXPERIMENTAL COMPARISONS

This section deals with the assessment of the effectiveness of the Bayesian approaches to the analysis/thresholding of the difference image proposed by the authors and described in Section 4. In particular, both the approaches to the estimation of the statistical terms of the difference image and the related Bayesian thresholding strategies are considered. Experiments were carried out on two different remote-sensing data sets.

The first data set consisted of two multispectral images acquired by the Landsat-5 Thematic Mapper (TM) sensor on the Island of Elba, Italy, in August and September 1994. Figures 4 (a) and (b) show channels 4 (i.e., near-infrared spectral channels) of both images.

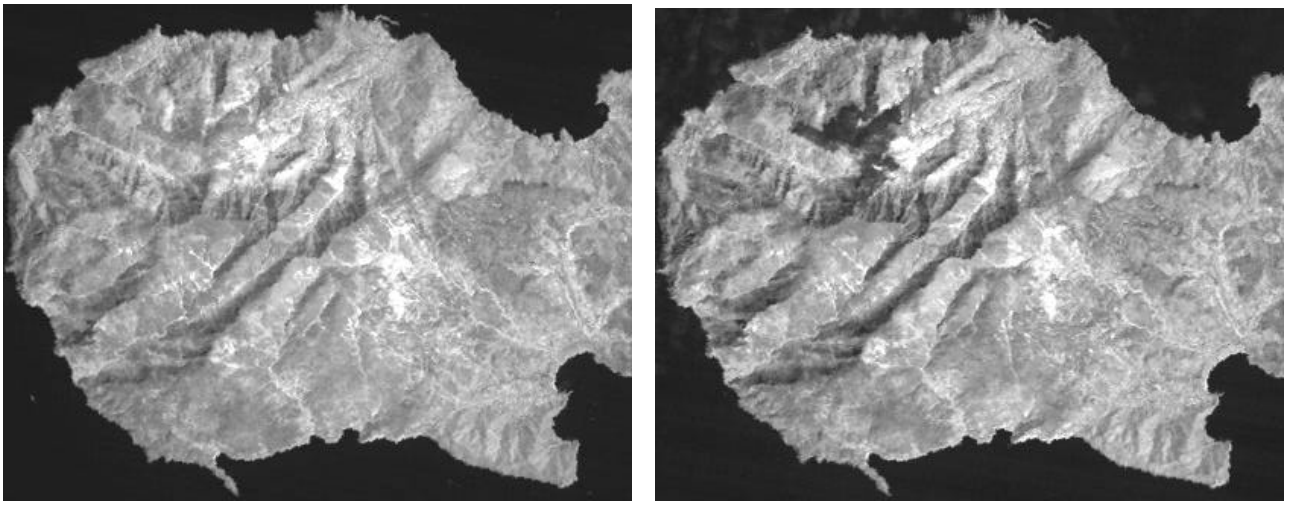
The second data set was composed of two images acquired on the Peloponnesian Peninsula, Greece, by the IRS-P3 WiFS sensor in April and September 1998. Figures 5 (a) and (b) show channels 2 (i.e. near-infrared spectral channels) of both images.

In both areas considered, some wildfires destroyed part of the vegetation between the two acquisition dates. The available ground-truth information about the burnt areas was used to prepare a "reference map" for each of the two data sets (Fig. 6 (a) and (b)). The reference maps turned out to be very useful to quantitatively assess change-detection errors. It is worth noting that this quantitative evaluation is a crucial task for algorithm validation.

Different experiments were carried out on the two data sets. The images of the Island of Elba were used to assess both the effectiveness of the parametric version of the unsupervised estimation technique described in Section 4.1 and the change-detection results obtained by the different non-contextual Bayesian strategies described in Section 4.2 (i.e., the minimum-error, minimum-cost, Neyman-Pearson, and minimax strategies). Concerning the images of the Peloponnesian Peninsula, they were used to validate both the adaptive semi-parametric version of the unsupervised estimation approach described in Section 4.1 and the spatial contextual strategy based on MRFs described in Section 4.2. In both cases, the results were compared with the ones obtained by applying to the difference image the optimal threshold T_o capable to minimize the overall error. Such a threshold was derived by a manual trial-and-error procedure according to the available reference maps.

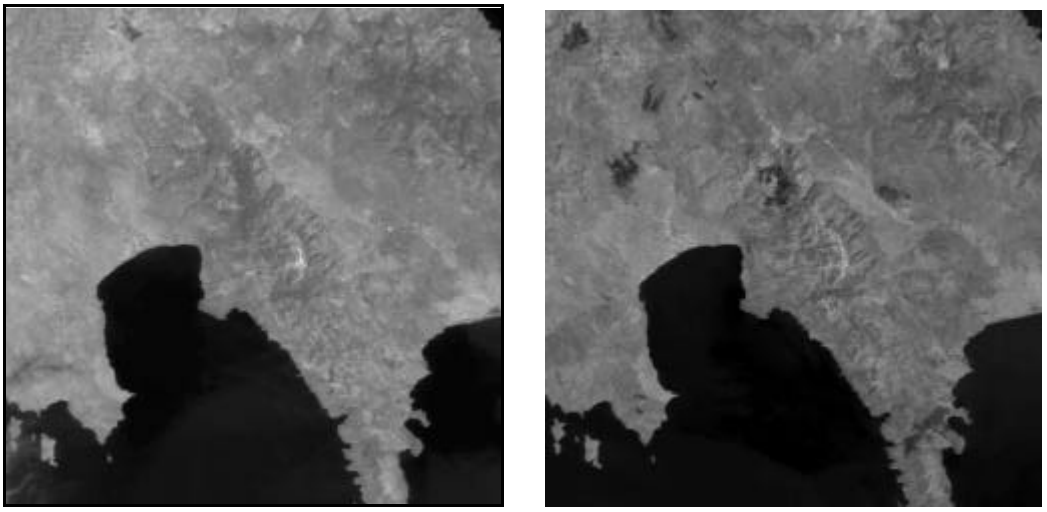
5.1 Data set related to the Island of Elba

As described in the methodological part of this chapter, the first step in a change-detection process is aimed at rendering the two multitemporal images comparable in both the spatial and spectral domains. To this end, the September 1994 image was registered to the August 1994 one. Concerning the radiometric calibration of the data, the analysis of the histograms of both images did not reveal any significant differences in the atmospheric and light conditions at the



(a) (b)

Fig. 4. Channels 4 of the Landsat TM images acquired on the Island of Elba, Italy, in: (a) August 1994; (b) September 1994.



(a) (b)

Fig. 5. Bands 2 of the IRS-P3 WiFS images acquired on the Peloponnesian Peninsula, Greece, in: (a) April 1998; (b) September1998.

(a) (b)

Fig. 6. Ground-truth maps of the changed areas used as reference maps in the experiments: (a) Island of Elba; (b) Peloponnesian Peninsula.

two acquisition times. Accordingly, no correction algorithms were applied.

The CVA technique was applied to spectral channels 4 and 7 of the images, as they were found to be very effective in locating the burnt areas. First of all, the density functions $p(X/w_n)$ and $p(X/w_c)$, and the prior probabilities $P(w_n)$ and $P(w_c)$ were derived. An analysis of the histogram of the difference image (see Fig. 7) revealed that it is quite reasonable to model the histogram as a mixture of two Gaussian functions (i.e., both density functions $p(X/w_n)$ and $p(X/w_c)$ can be approximated by a Gaussian function). Consequently, the parametric version of the approach described in subsection 4.1 was applied. In Fig. 7, the obtained estimates of the distributions of the two classes are superimposed upon the histogram of the difference image. As one can see, the density functions of the classes derived by the EM algorithm resulted in a reasonable approximation for the distribution of the difference image. Deeper quantitative analyses of the accuracy of the estimates and of the stability of the EM algorithm versus its initialization parameters were also carried out. The obtained results (reported in [24]) confirm the effectiveness of the estimation technique.

Once the estimates of the densities functions and of the prior probabilities of the classes were obtained, some of the decision strategies described in subsection 4.2 (i.e., the minimum-error, minimum-cost, Neyman-Pearson, minimax strategies) were applied to the difference image. The results were compared with the ones obtained by applying the threshold T_o that minimizes the overall error (i.e., 84). Such a threshold value was derived by applying a manual trial-and-error procedure to the difference image according to the ground-truth map.

First of all, the minimum-error criterion was used to derive the decision threshold (see (9)). A value of the threshold equal to 82 was obtained, which is very close to the minimum-error threshold T_o derived by the manual trial-and-error method (i.e., 84). Consequently, the automatic minimum-error strategy resulted in an overall change-detection error (i.e., 438 pixels) very close to the minimum one (i.e., 424 pixels). Concerning the error typologies, the proposed approach resulted in 218 false alarms and 220 missed alarms, whereas the minimum overall error threshold T_o involved 142 false alarms and 282 missed alarms. Figures 8 (a) and (b) show the change-detection maps obtained by applying the minimum error threshold T_o and the estimated threshold $\hat{T}_o$, respectively. As one can see, also a qualitative analysis of the obtained maps confirms the effectiveness of the proposed approach.

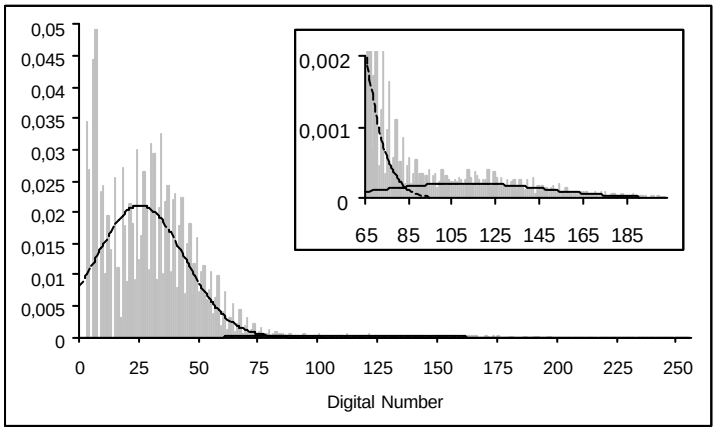
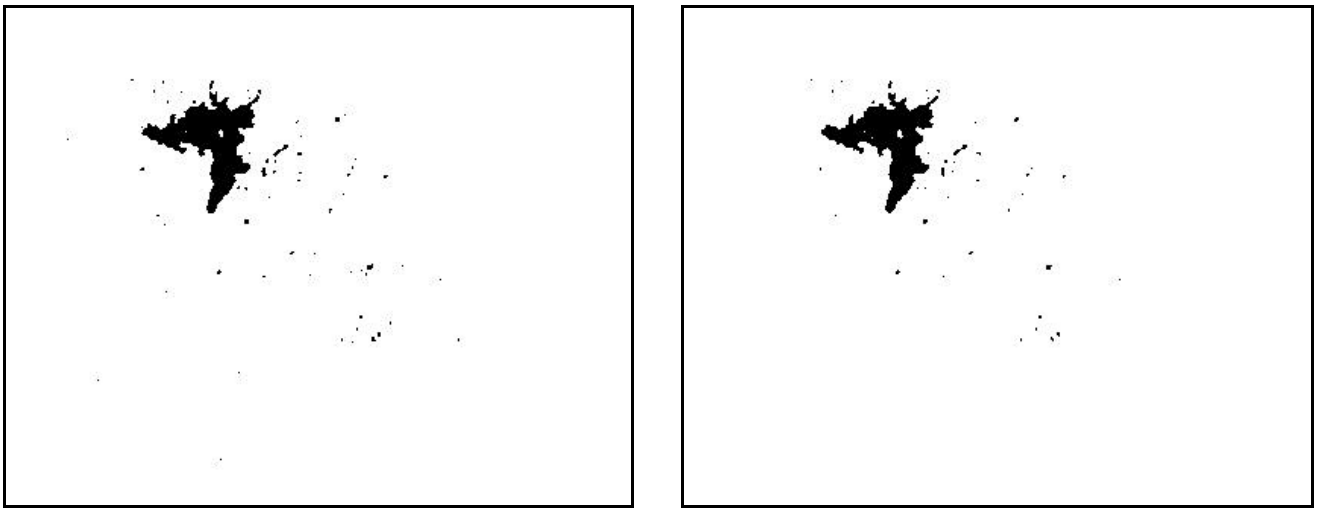


Fig. 7. Histogram of the difference image corresponding to the data set related to the Island of Elba. The estimates of the densities of the classes w_n and w_c, obtained with the parametric technique, are superimposed.

Let us now consider the case in which the costs of false and missed alarms are different and can be quantitatively evaluated. In this case, the decision threshold can be selected according to the minimum-cost criterion (see (10) and (11)). Two different cases may occur: i) end-users consider the underestimation of a burnt area (i.e., missed alarms) as the most critical error; ii) end-users require a final change-detection map where areas erroneously identified as burnt ones (i.e., false alarms) are minimized. In the first case, the cost c_{nc} associated with missed alarms was assumed k times as high as the cost c_{cn} associated with false alarms. In particular, two experiments with k equal to 5 and 10 were carried out. As one can see in Table 1, by increasing k, we significantly reduced the number of missed alarms (i.e., the more critical errors) from 151 (for k=5) to only 59 (k=10). As expected, the increase in k resulted in a larger number of false alarms. In the second case, the cost c_{nc} associated with missed alarms was assumed k times as low as the cost c_{cn} associated with false alarms. In particular, two experiments with k equal to 0.1 and 0.2 were carried out. Accordingly, the decrease in k resulted in a smaller number of false alarms (i.e., the more critical errors), and in a larger number of missed alarms (see Table 1). In both cases, the results meet the end-user's requirements, in agreement with the specific costs chosen. As an example, Fig. 9 shows the change-detection map obtained by applying the threshold estimated for k=0.2 (i.e., $\hat{T}_o = 92$). As one can see, the change-detection map, as compared with the one achieved by the minimum-error strategy (see Fig. 8 (a)), presents a smaller number of false alarms and a larger number of missed alarms.

Let us now suppose that end users require a specific constraint on the conditional probability of false alarms P_f. Table 2 gives the results obtained by applying the Neyman-Pearson criterion to the selection of the decision threshold for different values of P_f. As one can see, by decreasing the value of the acceptable P_f, we significantly reduced the number of false alarms from 1667 (P_f = 0.01) to 35 (P_f = 0.0001). On the contrary, the number of missed alarms increased, as expected.

Finally, let us consider the case in which the minimax criterion is adopted to derive the decision threshold. In this situation, the threshold is computed according to (14) for minimizing the maximum possible error. If such a criterion (for k=1) is applied, the estimated threshold is equal to 63. This value results in 3492 false alarms and 4 missed alarms.



(a) (b)

Fig. 8. Change-detection maps obtained for the data set related to the Island of Elba by applying to the difference image (a) the threshold estimated by the minimum-error Bayesian strategy, (b) the optimal threshold that minimizes the overall change-detection error yielded by a manual trial-and-error procedure.

Table 1. False and missed alarms incurred by the minimum-cost criterion for k values ranging from 0.1 to 10.

$K = c_{nc}/c_{cn}$	Threshold obtained	False Alarms	Missed Alarms
0.1	95	32	526
0.2	92	54	443
5	78	367	151
10	74	846	59

Table 2. False and missed alarms incurred by the Neyman-Pearson criterion for different constraints on the conditional probability of false alarms P_f.

P_f	Threshold obtained	False Alarms	Missed Alarms
0.0001	94	35	509
0.0005	87	122	311
0.001	83	218	220
0.005	73	871	53
0.001	69	1667	16

On the one hand, the use of the minimax criterion significantly increased the overall error (i.e., 3496 pixels), as compared with the case in which the prior probabilities of the classes are exactly known (i.e., 424 pixels); on the other hand, it decreased the overall error, as compared with the case in which the prior probabilities are unknown and significantly wrong estimates of such probabilities are used in the design of the decision rule according to (9) (e.g., using $P(w_n) = 0.1$ in the design of the decision rule, the overall error results in 10107 pixels).

5.2 Data set related to the Peloponnesian Peninsula

In this case too, the images were registered by using the image acquired in April as the reference one. The analysis of the histograms of both images did not reveal any significant differences in the atmospheric and light conditions at the two acquisition times. Consequently, no correction algorithms were applied. The UID technique was applied to spectral channel 2, as this band was found to be very effective in detecting burnt areas. The analysis of the histogram of the difference image (see Fig. 10) showed that it is not reasonable to model the density functions of the classes associated with changed and unchanged pixels with Gaussian functions. Consequently, the adaptive semi-parametric version of the unsupervised estimation approach was applied to derive $p(X/w_n)$, $p(X/w_c)$, $P(w_n)$, and $P(w_c)$. Concerning the initial thresholds, T_n and T_c, they were selected by setting a to 0.5 (see (2)); the number N of kernels and the smoothing parameter h were set to 6 and 50, respectively. In Fig. 10, the obtained estimates of the density functions of the classes w_n and w_c are superimposed

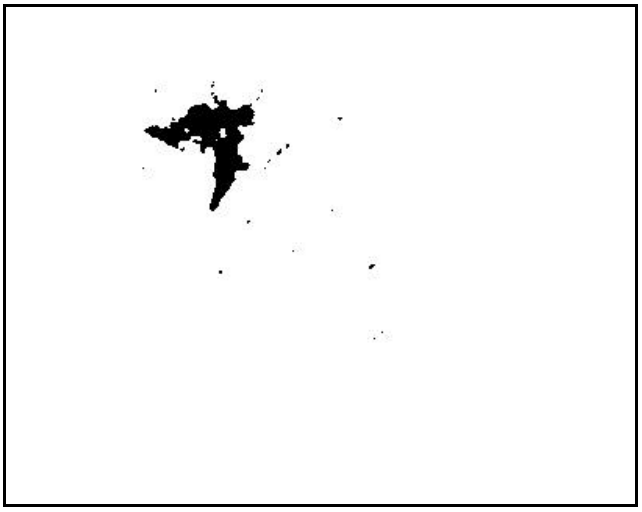


Fig. 9. Change-detection map obtained for the data set related to the Island of Elba by applying the threshold yielded by the minimum-cost strategy (k=0.2).

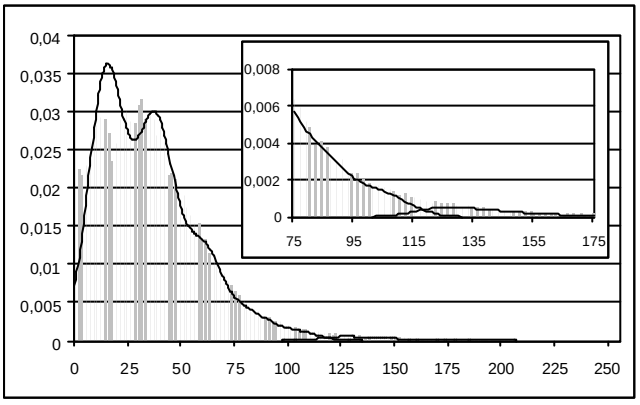


Fig. 10. Histogram of the difference image corresponding to the data set related to the Peloponnesian Peninsula. The estimates of the densities of the two classes w_n and w_c, obtained by the adaptive semi-parametric approach, are superimposed.

upon the histogram of the difference image. As one can see, such estimates provide an accurate description of the behavior of the histogram.

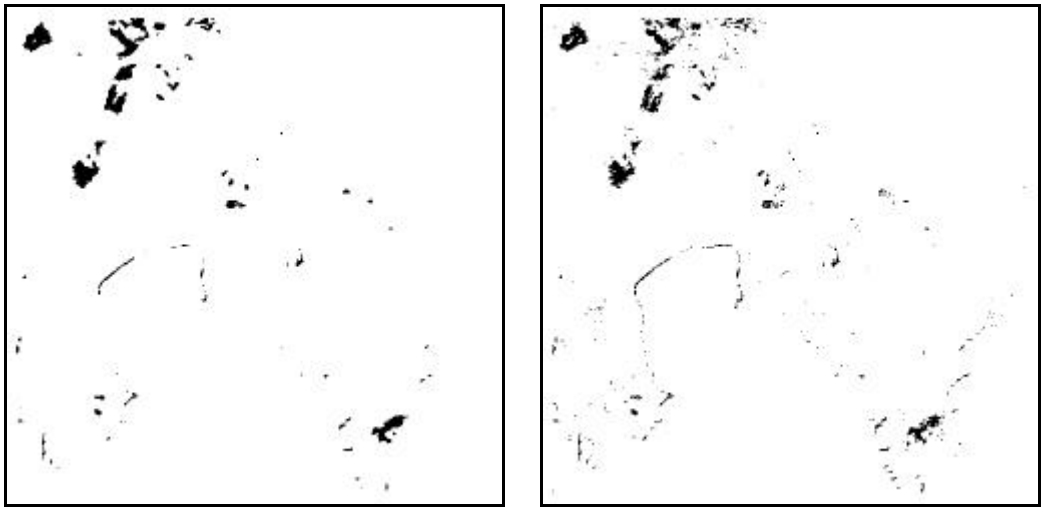
The aforementioned estimates can be exploited to derive the decision threshold by using any of the Bayesian strategies described in subsection 4.2. However, in order to provide an example of the results that can be obtained by using the context-spatial information in the analysis of the difference image, the strategy based on MRFs, as proposed in subsection 4.1.5, was applied to the difference image. The parameter b of the MRFs was set to 1.5 [25]. As one can see from Table 3, the overall error automatically obtained by the Bayesian contextual approach (i.e., 2763 pixels) was significantly smaller than the minimum one achievable by a manual trial-and-error procedure (i.e., 3553 pixels). In greater detail, on the one hand, the number of missed alarms was reduced from 1129 to 1010 pixels; on the other hand, the number of false alarms decreased from 2424 to 1753 pixels. A better understanding of the results can be achieved by analyzing of the change-detection maps obtained. In particular, in Fig. 11 the change-detection map obtained by exploiting the spatial context is compared with the one provided by the considered manual trial-and-error approach. A qualitative analysis of the two maps confirms that the use of the spatial contextual information allowed a more precise identification of the changed areas and yielded less noisy results.

6. CONCLUSIONS

In this chapter, different approaches and strategies have been presented that can be used to analyze/threshold the difference image in unsupervised change detection in multitemporal remote-sensing images. In particular, both heuristic approaches generally adopted in remote-sensing problems and Bayesian approaches recently proposed by the authors have been considered. In addition, several strategies for deriving the decision threshold according to the Bayesian decision theory have been proposed and discussed.

The effectiveness of one approach over another depends on the specific problem considered. Generally, heuristic trial-and-error approaches are very simple and do not require the development of advanced software modules. However, they exhibit several drawbacks: dependence of change-detection results on subjective evaluation criteria; low reliability; impossibility of realizing a completely automatic system, due to the need for the intervention of a human operator.

Concerning automatic heuristic strategies (e.g., methods based on statistical models, methods based on the optimization of an objective function of the selected threshold, etc.), although they overcome some of the aforementioned drawbacks, their effectiveness depends on the specific behavior of the histogram of the considered difference image. In addition, they are suitable for deriving change-detection maps with the objective of minimizing the



(a) (b)

Fig. 11. Change-detection maps obtained for the data set related to the Peloponnesian peninsula by applying (a) the context-based Bayesian strategy, (b) the optimal threshold that minimizes the overall change-detection error yielded by a manual trial-and-error procedure.

Table 3. Overall error, false and missed alarms related to the Peloponnesian peninsula data set. The results obtained by applying the automatic context-based strategy (context based) and the minimum-error threshold yielded by a manual trial-and-error procedure (minimum error) are given.

	False Alarms	Missed Alarms	Overall Error
Context based	1010	1753	2763
Minimum error	1129	2424	3553

overall error, but they are not effective if the objective of end-users is to minimize the overall change-detection cost or to perform change detection by imposing constraints on false or missed alarms. In addition, their empirical basis does not allow one to establish a relation with the theoretically optimal Bayes decision theory.

Bayesian strategies overcome the aforementioned problems by adopting different ways of defining the objective of change detection according to end-users' requirements (i.e., minimization of the overall error, minimization of the overall cost, constraints on the number of missed/false alarms). In addition, they are theoretically well-founded. However, the results obtained by these strategies strongly depend on the effectiveness of the unsupervised approaches used to estimate the prior probabilities and the probability density functions of changed and unchanged classes in the difference image. For this reason, we recommend both to accurately select the kind of approach to be used for the estimation process (i.e., parametric or adaptive semi-parametric) and to make a careful choice of the parameters to be given as input to the algorithms [24], [25]. It is worth noting that, when possible, the use of the parametric approach is to be preferred, as it requires the estimation of a smaller number of parameters, thus involving a more robust and more reliable estimation process. However, when the formulation of a reasonable assumption (or approximation) about the distribution of classes in the difference image is not possible, the adaptive semi-parametric method should be chosen.

Concerning the Bayesian strategies to be adopted for the choice of the threshold (i.e. the minimum-error, minimum-cost, Neyman-Pearson, and the context-based strategy), they should be selected on the basis of the available information in close cooperation with end-users, who well realize the consequences of each kind of error, and can therefore give useful suggestions. It is worth noting that, if changes to be identified are large enough as compared to the sensor geometrical resolution, we recommend to analyze the difference image by adopting the proposed spatial-contextual strategy based on MRFs. In fact, in this case, an efficient use of the interpixel class dependence context yields more reliable and more accurate change-detection results.

As a final remark, it is important to point out that, in the cases in which the decision threshold is defined in the design phase of the system and is then adopted also to analyze new images of the area considered, the use of the minimax criterion is recommended. This criterion allows one to avoid a sharp increase in the global error resulting from significant changes in the prior probabilities of changed and unchanged classes over time. In any case, if possible, it is more effective to automatically derive the decision threshold for each new pair of images considered, as the minimax criterion generally incurs much higher change-detection errors than the optimal Bayes rule for the minimum error.

ACKNOWLEDGMENTS

This research was supported by the Italian Space Agency (ASI). The authors wish to thank the Space Applications Institute (SAI) of the Joint Research Center (JRC) for providing the multitemporal images of the Peloponnesian Peninsula.

REFERENCES

[1] Goodenough, D. 2001, "Comparison of Methods for estimation of Kyoto Protocol Products of Forests from Multitemporal Landsat," *Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Synposium (IGARSS2001)*, 9-13 July 2001, Sydney, Australia.

[2] Singh, A. 1989, "Digital change detection techniques using remotely-sensed data," *International Journal of Remote Sensing*, 10, 6, 989-1003.

[3] Fung, T. 1990 "An assessment of TM imagery for land-cover change detection," *IEEE Transactions on Geoscience and Remote Sensing*, 28, 12, 681-684.

[4] Stauffer, M. L., and McKinney, R. L. 1978, *Landsat Image Differencing as an Automated Land Cover Change Detection Technique*, Technical memorandum Rep CSC/TM-78/6215, Computer Science Corporation, Silver Springs, Maryland, USA.

[5] Merril K. R., and Jiajun, L. 1998, "A comparison of four algorithms for change detection in an urban environment,"*Remote Sensing of the Environment*, 63, 95-100.

[6] Gopal S., and Woodcock, C. 1996, "Remote sensing of forest change using artificial neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, 34, 398-404.

[7] Fung, T., and LeDrew, E. 1988, "The determination of optimal threshold levels for change detection using various accuracy indices," *Photogrammetric Engineering & Remote Sensing*, 54, 10, 1449-1454.

[8] Singh, A. 1984, *Tropical forest monitoring using digital Landsat data in northeastern India*, Ph.D. thesis, University of Reading, Reading, England.

[9] Townshend, J. R. G., Justice, C. O., and Gurney, C. 1992, "The impact of misregistration on change detection," *IEEE Trans. on Geoscience and Remote Sensing*, 30, 5, 1054-1060.

[10] Flusser J., and Suk, T. 1994, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Trans. on Geoscience and Remote Sensing*, 32, 2, 382-387.

[11] Barnea, D. I., and Silverman, H. F. 1972 , "A class of algorithms for fast digital image registration," *IEEE Trans. on Computers*, C-21, 2, 179-186.

[12] Ton, J. and Jain, A. K 1989 "Registering Landsat images by point matching," *IEEE Trans. on Geoscience and Remote Sensing*, 27, 5, 642-650.

[13] Bruzzone, L., and Serpico, S. B. 1997 "Detection of changes in remotely sensed images by the selective use of multi-spectral information," *Int. Journal of Remote Sensing*, 18, 18, 3883-3888.

[14] Knoll, T. F., and Delp, E. J. 1986, "Adaptive gray scale mapping to reduce registration noise in difference images," *Computer Vision, Graphics, and Image Processing*, 33, 129-137.

[15] Gong, P., Ledrew, E. F., and Miller, J. R. 1992, "Registration-noise reduction in difference images for change detection," *Int. Journal of Remote Sensing*, 13, 4, 773-779.

[16] Chavez, P. S. 1989, "Radiometric calibration of Landsat Thematic Mapper multispectral images," *Photogrammetric Engineering & Remote Sensing*, 55, 9, 1285-1294.

[17] Chavez, P. S., and MacKinnon, D. J. 1994, "Automatic detection of vegetation changes in the southwestern United States using remotely sensed images," *Photogrammetric Engineering & Remote Sensing*, 60, 5, 1285-1294.

[18] Richards, J. A. 1993, *Remote Sensing Digital Image Analysis*, 2nd ed., Springer-Verlag, New York.

[19] Slater, P. N. 1987 "Reflectance and radiance based methods for the in-flight absolute calibration of multispectral sensors," *Remote Sensing of Environment*, 22, 11-37.

[20] Teillet, P. M., Slater, P. N., Ding, Y., Santer, R. P., R. Jackson, D., and Moran, M. S. 1990, "Three methods for the absolute calibration of the NOAA AVHRR sensors in flight," *Remote Sensing of Environment*, 31, 105-120.

[21] Olsson, H. 1995, "Reflectance calibration of thematic mapper for forest change detection," *Int. Journal of Remote Sensing*, 10.

[22] Rosin, P. L., "Unimodal thresholding," *Patter Recognition*, in press.

[23] Coppin, P., Jonckheere, I., Muys, B., Lambin, E. "Digital change detection methods in natural ecosystem monitoring: a review," *Proceedings of the First International Workshop on the Analysis of Multitemporal Remote-Sensing Images (Multitemp2001)*, 13-14 September 2001, Trento, Italy, in press.

[24] Bruzzone, L., Fernandez Prieto, D. 2000, "Automatic Analysis of the Difference Image for Unsupervised Change Detection," *IEEE Trans. on Geoscience and Remote Sensing*, 38, 3, 1171-1182.

[25] Bruzzone, L., Fernandez Prieto, D. 2001, "An Adaptive Semi-Parametric and Context-Based Approach to Unsupervised Change Detection in Multitemporal Remote-Sensing Images," *IEEE Transactions on Image Processing*, in press.

[26] Fukunaga, K. 1990, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, London.

[27] Stauffer, M. L., and McKinney, R. L. 1978, *Landsat image differencing as an automated land cover change detection technique*, Technical Memorandum Rep. CSC/TM-78/6215, Computer

Science Corporation, Silver Springs, Maryland, USA.

[28] Nelson, R. F. 1983, “Detecting forest canopy change due to insect activity using Landsat MSS,” *Photogrammetric Engineering & Remote Sensing*, 49, 1303-1314.

[29] O'Gorman, L. 1994, "Binarization and multi-thresholding of document image using connectivity," *CVGIP: Graphical Models and Image Processing*, 56, 494-506.

[30] Rosin, P.L., Ellis, T. J. 1995, “Image difference threshold strategies and shadow detection,” *British Machine Vision Conference*, 347-356.

[31] Otsu, N. 1979, “A threshold selection method from gray-level histogram,” *IEEE Trans. on Syst., Man., Cybern.*, 9, 62-66.

[32] Kapur, J.N., Sahoo, P.K., and Wong, A.K.C. 1985, "A new method for gray-level picture thresholding using entropy of the histogram," *Computer Vision, Graphics and Image Processing*, 29, 3, 273-285.

[33] Huang, L. K., and Wang, M. J. 1995, "Image thresholding by minimizing the measures of fuzziness," *Pattern Recognition*, 28, 41-51.

[34] Kittler, J. and Illingworth, J. 1986, "Minimum error thresholding," *Pattern Recognition*, 19, 41-47.

[35] Dempster, A. P., Laird, N. M., and Rubin, D. B. 1997, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of Royal Statistic. Soc.*, 39, 1, 1-38.

[36] Moon, T. K. 1996, “The Expectation-Maximization algorithm,” *Signal Processing Magazine*, vol.13, no.6, 47-60.

[37] Redner, A. P., and Walker, H. F. 1984, “Mixture densities, maximum likelihood and the EM algorithm,” *SIAM Review*, 26, 2, 195-239.

[38] Landgrebe, D. A. 1980, “The development of a spectral-spatial classifier for earth observational data,” *Pattern Recognition*, 12, 165-175.

[39] Priebe, C. E. and Marchette, D. J. 1993, “Adaptive mixture density estimation,” *Pattern Recognition*, 26, 5, 771-785.

[40] Fukunaga, K., and Hayes, R. 1989, “The reduced Parzen classifier,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 4.

[41] Duda, R. O., Hart, P. E., and Stork D. G., *Pattern classification*, 2nd ed., John Wiley & Sons, Inc, New York.

[42] Bruzzone, L. 2000, “An approach based on the Bayes rule for minimum cost to feature selection and classification of remote-sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, 38, 1.

[43] Bruzzone, L., Fernandez Prieto, D. 2000, "A minimum-cost thresholding technique for unsupervised change detection," *International Journal on Remote-Sensing*, 21, 18, 3539-3544

[44] Chellapa, R., and Jain A., Eds. 1993, *Markov Random Fields: Theory and Applications*. Academic, New York.

[45] Dubes, R. C., and Jain, A. K. 1989, “Random field models in image analysis,” *J. Appl. Statist.*, 16, 131-163, 1989.

[46] Rignot, E. and Chellappa, R. 1992, “Segmentation of polarimetric synthetic aperture radar data,” *IEEE Transactions on Image Processing*, 1, 3, 281-300.

[47] Smits, P. C., and Dellepiane, S. G. 1997, “Synthetic aperture radar image segmentation by a Detail Preserving Markov Random Field Approach,” *IEEE Transactions on Geoscience and Remote Sensing*, 35, 4, 844-857.

[48] Lakshmanan, S., and Derin, H. 1989, “Simultaneous parameter estimation and segmentation of Gibbs Random Fields using Simulated Annealing,” *IEEE Transactions on Geoscience and Remote Sensing*, 11, 8, 799-813.

[49] Chalmond, B. 1989, “An iterative Gibbsian technique for reconstruction of m-ary images,” *Pattern Recognition*, 22, 6, 747-762.

[50] Besag, J. 1986,“On the statistical analysis of dirty pictures,” *J. Royal Statist. Soc.*, ser. B, 48, 259-302.